

# Chapter 1

# Object Oriented Programming

## 1.1 Object oriented approach

### Definition 1.1.1: Object Oriented Programming

Object oriented programming (OOP) is a imperative programming paradigm where instructions are grouped together with the part of the state they operate on.

The main **structural components** of all systems are

- Objects: Object is something that takes up space in the real or conceptual world with which somebody may do things. An object is an instance of a class. An object has three main characteristics:
  - Name (or ID): Unique identifier for the object.
  - State: Attributes that describe the object.
  - Operations (or behavior): Actions that the object can perform.
- Class objects: Class is the blueprint of an object.

Main **characteristics** of the approach are

- Encapsulation: The mechanism of hiding the implementation of the object. Only the necessary details are exposed to the user.
- Abstraction: A principle which consists of ignoring the aspects of a subject that is not relevant for the present purpose. It is the process of simplifying complex systems by modeling classes based on the essential properties and behaviors an object must have.
- Inheritance: Mechanism by which one class can inherit attributes and methods from another class. Making new classes from existing ones.
- Polymorphism: Ability to present the same interface for different data types. Same function name but different signatures being used for different types.

## 1.2 Class

### Definition 1.2.1: Class Diagram

A class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

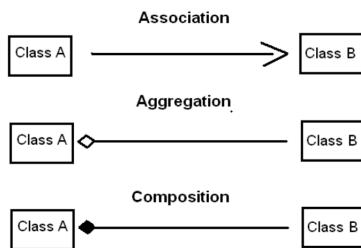
Elements of a **class**:

- Class Name (ID Class): The name of the class. Nouns associated with the textual description of a problem. Singular.

- Attributes: The properties or characteristics of the class. Types include Real, Integer, Text, Boolean, Enumerated, etc.
- Operations (Methods): The functions or operations that can be performed on the class. Behaviors of the class.

### Relations between classes:

- Association: A relationship between two classes that indicates how objects of one class are connected to objects of another class.
- Aggregation: A special type of association that represents a "whole-part" relationship between classes.
- Composition: A stronger form of aggregation that implies ownership and a whole-part relationship where the part cannot exist independently of the whole.
- Generalization: A relationship between a more general class (superclass) and a more specific class (subclass) that indicates inheritance.



### 1.2.1 Class and Method in Python

```

1   class Person:
2       pass # An empty class definition
3
4   p = Person()
5   print(p) # Output: <__main__.Person object at 0x...>

```

Listing 1.1: Empty class definition in Python

```

1   """Define class with method"""
2   class Person:
3       def speak(self):
4           print("Hello!")
5
6   """Create object and call method"""
7   p = Person()
8   p.speak() # Output: Hello!

```

Listing 1.2: Method

```

1   """
2   The __init__ method is a special method in Python classes.
3   It is a method that Python calls when you create a new instance of this class.
4   """
5   class Person:
6       def __init__(self, name):
7           self.name = name # Initialize the name attribute
8       def speak(self):
9           print('Hello, my name is', self.name)
10
11 p = Person('Carlos')
12 p.speak() # Output: Hello, my name is Carlos

```

Listing 1.3: Attributes and `__init__` method

**Note:**

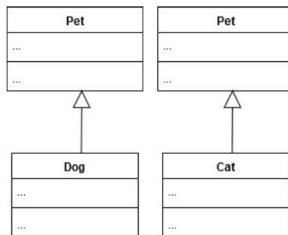
The first argument of every class method, including `init`, is always a reference to the current instance of the class. By convention, this argument is always named `self`.

```
1  class Pet(object):
2      def __init__(self, name, species):
3          self.name = name
4          self.species = species
5      def getName(self):
6          return self.name
7      def getSpecies(self):
8          return self.species
9      def __str__(self):
10         return "%s is a %s" % (self.name, self.species)
```

Listing 1.4: Class Pet with attributes and methods

```
1  class Dog(Pet):
2
3      """Dog inherits from Pet"""
4
5      def __init__(self, name, chases_cats):
6          # Call the parent class (Pet) constructor
7          Pet.__init__(self, name, "Dog")
8          self.chases_cats = chases_cats
9
10     def chasesCats(self):
11         return self.chases_cats
12
13 class Cat(Pet):
14
15     """Cat inherits from Pet"""
16
17     def __init__(self, name, hates_dogs):
18         # Call the parent class (Pet) constructor
19         Pet.__init__(self, name, "Cat")
20         self.hates_dogs = hates_dogs
21
22     def hatesDogs(self):
23         return self.hates_dogs
```

Listing 1.5: Inheritance



**Example 1.2.1 (Inheritance)**

```
myPet = Pet("Boby", "Dog")
myDog = Dog("Rex", True)
```

- `isinstance(myDog, Pet)` # Returns: True
- `isinstance(myDog, Dog)` # Returns: True

- `isinstance(myPet, Pet)` # Returns: True
- `isinstance(myPet, Dog)` # Returns: False

### 1.2.2 Access Modifiers

There are three types of access modifiers in Python:

- Public: Members (attributes and methods) declared as public are accessible from anywhere.
- Protected: Members declared as protected are accessible within the class and its subclasses. In Python, this is indicated by a single underscore prefix (e.g., `_attribute`).
- Private: Members declared as private are accessible only within the class itself. In Python, this is indicated by a double underscore prefix (e.g., `__attribute`).

```

1   class Person:
2       def __init__(self, name, age):
3           self.__name = name      # Private attribute
4           self.__age = age        # Private attribute
5
6   p = Person("David", 23)
7   p.__name # This will raise an AttributeError

```

Listing 1.6: Private Attributes

```

1   class Person:
2       def __init__(self, name, age):
3           self.name = name      # Public attribute
4           self.age = age        # Public attribute
5
6   p = Person("David", 23)
7   p.name # This will work fine and return "David"

```

Listing 1.7: Public Attributes

```

1   class Person:
2       def __init__(self, name, age):
3           self._name = name      # Protected attribute
4           self._age = age        # Protected attribute
5
6   p = Person("David", 23)
7   p._name # This will work, but it's discouraged to access protected members directly

```

Listing 1.8: Protected Attributes

```

1   class Person:
2       def __init__(self, money=0, energy=100):
3           self.money = money
4           self.energy = energy
5
6       def work(self, hours):
7           if self.energy >= hours * 10:
8               self.money += hours * 10 # Assume earning $10 per hour of work
9               self.energy -= hours * 10
10              print(f"Worked for {hours} hours. Money increased to ${self.money}, energy decreased to {self.energy}.")
11
12          else:
13              print("Not enough energy to work.")

```

Listing 1.9: Example Usage

# Chapter 2

## Data Analysis and Statistics

### 2.1 Numpy

#### Definition 2.1.1: Numpy

Numpy is a fundamental package for scientific computing in Python. It provides support for arrays, matrices, and many mathematical functions to operate on these data structures efficiently.

#### 2.1.1 Shape, rank, and size

```
1 c = np.array([1, 2, 3, 4]) # Create a 1D array
2 print(type(c)) # Output: <class 'numpy.ndarray'>
3
4 b = np.array([[1, 2, 3], [4, 5, 6]]) # Create a 2D array
5 shape = b.shape # Output: 2, 3
6 rank = np.ndim(b) # Output: 2
7 size = b.size # Output: 6
```

Listing 2.1: Shape, rank, and size

#### 2.1.2 Manipulating arrays

**Reshaping arrays** means changing the shape of an array. The shape of an array is the number of elements in each dimension. By reshaping, we can add or remove dimensions or change number of elements in each dimension.

```
1 arr = np.array([4, 4, 3, 4, 5, 6, 7, 9, 0, 10, 17, 12])
2 newarr = arr.reshape(3, 4) # Reshape to 3 rows and 4 columns
3 print(newarr)
4 # Output:
5 # [[ 4  4  3  4]
6 #  [ 5  6  7  9]
7 #  [ 0 10 17 12]]
```

Listing 2.2: Reshaping an array

We can also access an array element and change value to an array.

```
1 a = np.array([10, 20, 30, 40, 50])
2 a[2] = 50 # Change the third element to 50
3 print(a) # Output: [10 20 50 40 50]
```

Listing 2.3: Accessing and changing array elements

```

1 a = np.zeros((2,2)) # Create a 2x2 array of zeros
2 b = np.ones((3,3)) # Create a 3x3 array of ones
3 c = np.full((2,3), 7) # Create a 2x3 array filled with 7s
4 d = np.eye(3) # Create a 3x3 identity matrix

```

Listing 2.4: Creating special arrays

**Joining NumPy arrays** means putting contents of two or more arrays into a single array.

```

1 a = np.array([1, 2, 3])
2 b = np.array([4, 5, 6])
3
4 c = np.concatenate((a, b)) # Join a and b
5 print(c) # Output: [1 2 3 4 5 6]
6
7 d = np.stack((a, b), axis=0) # Stack a and b vertically
8 print(d)
9 # Output:
# [[1 2 3]
10 # [4 5 6]]
11
12 e = np.stack((a, b), axis=1) # Stack a and b horizontally
13 print(e)
14 # Output:
# [[1 4]
15 # [2 5]
16 # [3 6]]
17
18

```

Listing 2.5: Joining NumPy arrays

```

1 a = np.array([[1,2,3,4], [5,6,7,8], [9,10,11,12]])
2 print(a)
3 # Output:
# [[1 2 3 4]
4 # [5 6 7 8]
5 # [9 10 11 12]]
6
7 b = a[:3, 1:3] # Slicing rows 0-2 and columns 1-2
8 print(b)
9 # Output:
# [[2 3]
10 # [6 7]
11 # [10 11]]
12
13

```

Listing 2.6: Slicing NumPy arrays

### Problem 2.1: Slicing arrays

b[0, 0] = 20

What happens to array a?

**Answer:** Array a is also changed because b is a view of a.

### Problem 2.2: Slicing arrays

How to get a copy of a slice of an array so that changes to the slice do not affect the original array?

**Answer:** Use the .copy() method to create a copy of the slice.

```

1 b = a[:3, 1:3].copy() # Create a copy of the slice
2 b[0, 0] = 20 # Change the first element of b

```

```

3     print(a) # Array a remains unchanged
4
5     # Output:
6     # [[ 1  2  3  4]
7     # [ 5  6  7  8]
8     # [ 9 10 11 12]]

```

Listing 2.7: Creating a copy of a slice

### 2.1.3 Handling indexes

```

1     a = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9], [10, 11, 12]])
2     print(a)
3     # Output:
4     # [[ 1  2  3]
5     # [ 4  5  6]
6     # [ 7  8  9]
7     # [10 11 12]]
8
9     b = np.array([0, 2, 0, 1])
10    print(a[np.arange(4), b]) # Access elements using array of indices
11    # Output: [ 1  6  7 11]
12
13    a[np.arange(4), b] += 10 # Increment selected elements by 10
14    print(a)
15    # Output:
16    # [[11  2  3]
17    # [ 4  5 16]
18    # [17  8  9]
19    # [10 21 12]]

```

Listing 2.8: Fancy indexing

```

1     b = 1
2     b += 10
3     print(b) # Output: 11
4
5     b = 1
6     b = b+10
7     print(b) # Output: 11

```

Listing 2.9: Boolean indexing

```

1     grades = np.array([14, 12, 13, 14, 15, 14, 14])
2     x = np.where(grades == 14) # Find indices where grades are 14
3     print(x) # Output: (array([0, 3, 5, 6]),)

```

Listing 2.10: Searching arrays

### 2.1.4 Random

#### Definition 2.1.2: Random

Random number does not mean a different number each time. Random means something that cannot be predicted logically.

```

1     from numpy import random
2
3     # Generating random numbers within an array
4     e = np.random.random((4, 4)) # Create a 4x4 array of random floats in [0.0, 1.0)

```

```

5      # Generating random numbers
6      x = random.randint(100) # Generate a random integer from 0 to 99
7
8      # Generating arrays of random numbers
9      x = random.randint(100, size=(5)) # Generate an array of 5 random integers from 0 to 99
10     x = random.randint(100, size=(3, 4)) # Generate a 3x4 array of random integers from 0
11         to 99
12     x = random.rand(5) # Generate an array of 5 random floats in [0.0, 1.0)

```

Listing 2.11: Generating random numbers

### Definition 2.1.3: Random data distribution

A random distribution is a set of random numbers that follow a certain probability density function (PDF).

- Normal distribution (`random.normal`)
  - loc: (mean) where the peak of the bell curve is located
  - scale: (standard deviation) how flat and wide the bell curve is
  - size: output shape
- Binomial distribution (`random.binomial`)
  - n: number of trials
  - p: probability of success on each trial
  - size: output shape

```

1      # Normal Distribution
2      x = random.normal(loc=0.0, scale=1.0, size=(2,3)) # Generate a 2x3 array of random
3          numbers from a normal distribution
4      print(x)
5      # Output:
6      # [[ 0.49671415 -0.1382643   0.64768854]
7      #   [ 1.52302986 -0.23415337 -0.23413696]]
8
9      # Binomial Distribution
10     x = random.binomial(n=10, p=0.5, size=(3,4)) # Generate a 3x4 array of random numbers
11         from a binomial distribution
12     print(x)
13     # Output:
14     # [[5 6 4 3]
15     #   [4 7 6 4]
16     #   [6 3 5 4]]

```

Listing 2.12: Generating random numbers from different distributions

## 2.2 Pandas

### Definition 2.2.1: DataFrame

Labelled data structure with columns of potentially different types. Like a spreadsheet or SQL table, or a dict of Series objects.

```

1   d = {
2     'col1': [1,2,1,3,1,2],
3     'col2': [1,2,3,4,5,6]
4   }
5   df = pd.DataFrame(data=d)
6
7   df.count() # Count non-NA cells for each column
8   df['col1'].value_counts() # Count unique values in col1
9   df['col1'][1] = 5 # Change value in col1 at index 1 to 5

```

Listing 2.13: Creating and manipulating a DataFrame

### Problem 2.3: Copy DataFrames

```

col1 = df['col1'] # Accessing a single column
col1[2] = 99
What is the result in col1 and df?
Answer: Both col1 and df are changed because col1 is a view of df.

```

### Problem 2.4: Copy DataFrames

```

new_col1 = col1.copy()
new_col[2] = 9999

```

What is the result in new\_col1 and df?  
**Answer:** Only new\_col1 is changed, df remains unchanged because new\_col1 is a copy of col1.

```

1   df = pd.read_csv('data.csv') # Read data from a CSV file into a DataFrame
2   df.to_csv('output.csv') # Write DataFrame to a CSV file

```

Listing 2.14: Read and save

```

1   df.head() # Display the first 5 rows of the DataFrame
2   df.info() # Display a summary of the DataFrame
3   df.describe() # Generate descriptive statistics
4   df.columns # List all column names

```

Listing 2.15: DataFrame information

```

1   df.at(2, 'col1') # Access a single value for a row/column label pair
2   df.iloc[2, 0] # Access a single value for a row/column pair by integer position
3   df.xs(2) # Returns cross-section (row) at index 2
4   df.loc[:, 'col1'] # Access a group of rows and columns by labels or a boolean array

```

Listing 2.16: Accessing DataFrame elements

### Example 2.2.1 (Access to row and columns)

- Cells: df.iloc[195][0]
- Rows: df.iloc[[195], :]
- Columns: df.loc[:, 'col1']

```

1   df1=df2 # Assignment
2   copydf = df.copy(deep=False) # Shallow copy
3   copydf = df.copy(deep=True) # Deep copy

```

---

Listing 2.17: Copying DataFrames

```
1 # Convert column to numeric, setting errors to NaN
2 df['GDP'] = pd.to_numeric(df['GDP'], errors='coerce')
3
4 # Suppose you have , instead of . in the numbers
5 df['GDP'] = df['GDP'].str.replace(',', '.',)
6
7 # Suppose you have $ or € symbols in the numbers
8 df['GDP'] = df['GDP'].str.replace('$', '', regex=True).str.replace('€', '', regex=True)
9
10 # Drop rows with any NaN values
11 df.dropna(inplace=True)
```

Listing 2.18: Data cleaning and conversion

```
1 X.mean() # Calculate mean of each column
2 X.median() # Calculate median of each column
3 X.mode() # Calculate mode of each column
4 X.std() # Calculate standard deviation of each column
5 X.var() # Calculate variance of each column
6 X.corr() # Calculate correlation matrix
7 X.cov() # Calculate covariance matrix
8 X.max() # Calculate maximum of each column
9 X.min() # Calculate minimum of each column
10 X.kurt() # Calculate kurtosis of each column
11 X.skew() # Calculate skewness of each column
```

Listing 2.19: Statistical methods

## 2.3 Regular Expressions

## 2.4 Statistics

## 2.5 Time Series

## 2.6 Regression